DOCUMENT RESUME

ED 034 303 EA 002 624

AUTHOR	Vincent, William S.
TITLE	Further Clarification of the Class Size Question.
INSTITUTION	Columbia Univ., New York, N.Y. Inst. of Administrative Research.
PUB DATE	Nov 68
NOTE	4p.
JOURNAL CIT	IAR-Research Bulletin; v9 n1 p1-3 Nov 68
EDRS PRICE	EDRS Price MF-$0.25 HC-$0.30
DESCRIPTORS	*Class Size, *Elementary Grades, *Measurement Instruments, *Research Reviews (Publications), *Secondary Grades

ABSTRACT

This research is directed at finding an optimum class size in elementary and secondary grades. The Indicators of Quality program observes selected classroom characteristics and scores each characteristic positive, zero, or negative. The difference between the positive and negative scores provides a measure to compare with class size. Data from 47 school districts of the Metropolitan School Study Council show a progressively larger difference as class size decreases. In the elementary grades a significant break occurs between the 11-15 and 16-20 and the 21-25 and 26-30 class size intervals. In the secondary grades, the only significant break occurs between the 11-15 and 16-20 class size intervals. This process measure provides useful categories on large and small classes. Combining the results of this study with achievement test criterion will further resolve the class size question. (LN)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

IAR- Research Bulletin
INSTITUTE OF ADMINISTRATIVE RESEARCH
Teachers College • Columbia University

Volume 9, No. 1 November, 1968

ED034303

Further Clarification of the Class Size Question

WILLIAM S. VINCENT

The first article in this **Bulletin** was a review of class size studies and a report of an analysis of staff adequacy in relation to expenditure level.[1] Since then two other articles have been devoted to class size—one by McKenna and Pugh[2] and one by Woodson.[3] Each contributed its bit toward illuminating this subject about which there has been so much discussion and, indeed, investigation without much clarification. Recently the most extensive class size study yet reported resulted as a by product of an entirely different kind of investigation. Actually the piece of information pertaining to class size is relatively minor. Together with what has gone before, however, one can begin to fit the pieces together into a semblance of that the final answers to the class size question will probably look like.

The class size question, once referred to by Ross and McKenna as the "million dollar question"[4] is briefly this: Do pupils learn better in smaller classes than they do in larger classes? To quote the earlier article, "It is obvious that the question has fiscal implications. Other things being equal, the school district that puts fewer pupils in each class spends more; the district bent on saving money may think it can do so by increasing class size."[1] To this simple question, however, there are two qualifiers. Does the additional cost of small classes result in a sufficient additional benefit in pupil learning, assuming there is any at all? Is the class size/pupil benefit relationship smooth and linear or is there — as seems more likely — some critical break point, such that change in class size above and below this optimum has little effect?

EA 002 624

The answer to the first of these is still unclear. We cannot say at this point what percentage of the variance in pupil accomplishment or pupil benefit, measured in any acceptable manner, is accounted for by variance in class size. Even if we were able to say, the question of whether the benefit were worth the cost is a subjective determination. How large is a dollar's worth of benefit? Would any degree of benefit, even a small increment in the final outcome of pupil achievement, be worth whatever it might cost in the conservation of human resources? Technicians cannot answer such questions; only those who set the objectives and policies for education can do so.

An answer to the second question, however, can be given. There does appear to be an optimum class size. There seems to be a point below which benefit is greatest above which change in class size does not result in increased or decreased benefit. This break point seems to be remarkably consistent for both elementary and secondary classes. Before examining the evidence, however, let us review what the previously mentioned articles had to say about class size.

THE TIME SCALE STUDY

The first one[1] was based on a study of 132 school districts in 33 states. The staffing practices of these districts were first viewed in relation to a general school quality criterion. This particular criterin was the **Time Scale;**[5] a simple device for measuring adaptability, or capacity for innovation, of a school district by determining the number of innovations present and the date at which each was adopted. The zero order correlation between class size and criterion was not high. But when the population of school districts was divided into high, middle and low expenditure groups, clear differences in staffing adequacy could be seen among them, since the quantity of staff employed per 1000 pupils is obviously an expenditure related variable. The relation between numberical staff

1. "The Question of Class Size," IAR Research Bulletin, Volume 1, No. 1, October, 1960.
2. "Performance of Pupils and Teachers in Small Classes Compared to Large," IAR Research Bulletin, Volume 4, No. 2, February, 1964.
3. "Effect of Class Size as Measured by an Achievement Test Criterion," IAR Research Bulletin, Volume 8, No. 2, February, 1968.
4. Donald H. Ross and Bernard McKenna, **Class Size: The Multi-Million Dollar Question,** New York, Metropolitan School Study Council, 1957.
5. Paul R. Mort and Truman M. Pierce, **A Time Scale for Measuring the Adaptability of School Systems,** New York, Metropolitan School Study Council, 1947.

adequacy (NSA: total number professionals per thousand pupils) and net current expenditure per pupil provides a regression equation by means of which a "staffing residual" was calculated for each district in the sample. This staffing residual is the difference between the actual NSA and that predicted by the correlation between NSA and expenditure. This residual was then related to the **Time Scale** score with a remarkable result. For middle expenditure districts the correlation was negative M indicating that better schools the middle expenditure range had **smaller** staffs in proportion to number of pupils. These schools were putting marginal dollars into teachers salaries rather than employing more teachers. For **both** the high and low expenditure groups, on the other hand, the correlation was positive. With respect to staffing, districts at opposite ends of the expenditure scale employed similar policies as these policies were related to school quality.

Thus the relation between class size and quality, which might appear at first to be confused or inconclusive, is a factor intimately bound up with fiscal policy. One cannot, it would seem, draw conclusions regarding the influence of class size policy without knowing something about fiscal policy. When this is done, sharp relationships stand out (the correlations referred to above were of the order of .46 to .58).

THE WOODSON STUDY

The next probe of overall district policy regarding class size was made by Marshall Woodson.[3] Here again his objective was not to measure point by point increments in benefit that may accompany point by point decrease in class size. Rather his class size variables were all intended to be district wide measures reflecting general class size policy — average class size of district, class size range, and percentage of classrooms with less than 22 pupils and percentage of classrooms with more than 27. The last two turned out to be the most effective indices of district policy and showed a consistent relationship to achievement test residuals. The criterion in this instance was the difference between actual and predicted achievement test scores, the prediction being based on the regression between IQ and achievement as reported by the achievement test publishers. Woodson found that for the arithmatic and reading subtests and the overall total test, correlations between the residuals computed for each district and percentage of classrooms in the district with more than 27 pupils were consistently negative, and that correlations between residuals and percentage of classrooms with less than 22 pupils were consistently positive, although significance at the .05 level appeared only for the low IQ group (IQ below 85).

He divided the districts into upper and lower thirds in the class size range and computed arithmetic, reading, and overall total scores for low, middle, and high ability groups and for the total group of both fourth and sixth grade pupils. He found that in these various comparisons the mean criterion scores, computed as standard scores of the residuals, were without exception higher for the group of districts in the lower third of the class size range. Similarly, he divided the districts into the upper third and lower third of the criterion scale and found that, in a similar set of comparisons, the mean class size of comparisons, the mean class size of districts in the lower third of the criterion range was in almost every instance larger than the mean class size of districts in the upper third. The highest levels of significance appeared in comparing districts on the basis of percent of classes with less than 22. Districts in the upper third of the criterion range had a greater percentage of classes less than 22 than districts in the lower third of the criterion range.

The assumption of a boundary of 22 for the upper limit of the "small class" and 27 for the lower limit of the "large class" was mostly suppositional on Woodson's part. An upper boundary of 20 had been the definition of the "small class" for the Metropolitan School Study Council Commission on the School of 1980. This too was mostly a guess. No one had attempted to establish the dimensions of largeness and smallness in dealing with class size. In fact, one of the weaknesses of the bulk of class size research is the fact that there is no agreement on the quantitative dimensions of the terms employed. The size range that in one study is viewed as "small" turns out to be "large" in some other study.

THE INDICATORS OF QUALITY INVESTIGATION OF CLASS SIZE

We now find that there is evidence to support a specific break point in the secondary and a series of two break points in the elementary levels that, among other things, specify the upper limit of "small class" or the lower limit of "large class." This information comes as a by product, as suggested above, of another investigation. This other investigation was the Spring, 1967, applications of **Indicators of Quality** in 47 school districts of the Metropolitan School study council. Indicators was applied in 4283 classrooms during this investigation—2106 in the third, fourth, fifth, and sixth grades, and 2181 in the tenth, eleventh, and twelfth grades. More on the outcome of this project, whose results are still being examined statistically, will be published during the year in the pages of the **Bulletin.** However, the piece of information on class size was an early outcome of computer runs.

The data are presented in the accompanying table. **Mean difference score** is one of the methods of scoring the results of **indicators** where the total negative signs seen are subtracted from the total positive signs seen. The higher the score the better. The mean difference scores shown for each class interval represent the mean of all the net scores so obtained in the classrooms observed in the various grades at that class size interval. It will be seen that the mean of each group winds up positive, although there would be individual cases whose difference score was negative.

It will be seen first of all that there is a relationship between class size and score. Since this score is obtained on the basis of observing events occurring in the classroom it is thus a **process measure.** This is the kind of criterion employed by Pugh.[6] Woodson's criterion, based on achievement test results, is classed as an **output criterion.**[7] Nevertheless, despite the differences in the criteria employed, the indications ar similar.

OPTIMUM CLASS SIZE

The principal message in the table, however, is the breakpoint between intervals 11-15 and 16-20. It is strong and definite at the secondary level, provisional at the elementary. A second breakpoint occurs in the elementary level between intervals 21-25 and 26-30. The differences at these points are significant at the .01 level, even at the sparcely populated lower end of the elementary scale

6. James B. Pugh, Jr., An Analys s of the Characteristics of Teaching and Learning Related to Pupil-Teacher Ratio, New York: nstitute of Administrative Resarch, 1964.

when all 119 cases below a class size of 16 are summed and the mean of the total compared to the next interval. (It should be noted that the score of 9.10 based on 10 cases at interval 36-40 is not significant.)

What can one conclude from this? One would say that the general parameters qualifying the class size question have begun to come clear. We now have a basis for distinguishing between "large" and "small" in class size. We can say that a class size study using some other breakpoint between "large" and "small" can only provide us with results that are inconsequential. And this may in part explain the generally inconclusive results of the corpus of class size investigations. In the secondary school "large" appears to be "sixteen pupils or more;" while "less than sixteen" appears to define the "small" class. In the elementary school there appear to be **three** differentiations of class size. The "very small" (less than sixteen), the "medium small" (sixteen to twenty-five), and the "large" (more than twenty five). Thus there is some justification for the "fewer than 22" category employed by Woodson and the "fewer than 20" employed by the Council Commission on the School of 1980. Nevertheless, differentiation at 20 or 22 clouds the distinction between "very small" and "medium small" at the elementary level, which we see to be significantly different in the criterion measure, and it is of course useless at the secondary level.

We find also that smaller classes are significantly better than larger classes when measured by this kind of a criterion. Since **indicators** is based upon events in the classroom that characterize individualization of instruction, interpersonal regard, creativity, and group activity, all areas of school process that authorities on learning average are critical to the learning process, one can conclude that more of this kind of teaching/learning activity does take place in classes whose numbers fall below the significant breakpoints. This tends to confirm the results of Woodson's study, (achievement test) and to suggest that if the more significant breakpoints had been employed by him his results would have been characterized by more significant differences. Moreover, it is apparent that manipulation of class size above and below these optimum points accomplishes little as measured by such a criterion. The difference between the criterion score at the 31-35 class size interval and the 26-30 interval, for example, is insignificant. Thus we see that the class size policy of the middle expenditure schools in the **Time Scale** study cited above was wisely adopted. Since their resources were not sufficient to make a large enough difference in class size, they elected to exert fiscal effort on staff improvement through salary policy. The high expenditure schools, on the other hand, possessed the fiscal resources to do both, reduce class size to something approaching significant levels and at the same time maintain an attractive salary policy.

We now have the basis for the ultimate probe. Both a process measure and achievement test criterion based, like the Woodson measure, on residuals should be applied to a stratified sample of classes at least as large as the one reported here. Differences between the two levels of size in the secondary grades and the three levels of size in the elementary grades should be computed using the boundaries distinguishing "large" and "small" revealed by this study. Scores expressing a degree of adherence to a class size policy of large or small should then be computed for each school district in the sample. These scores, together with the criterion scores, and measures of significant inputs - - finance, staff characteristics, staff deployment — should be fed into a multivariate program. From this it could be determined how much of the variance in the criteria is accounted for by all the inputs, including class size, and what proportion of its class size uniquely accounts for. This would settle the class size question.

7. "Measuring School Quality Input and Criteron," IAR Research Bulletin, Volume 6, No. 1, November, 1965.

TABLE

MEAN DIFFERENCE SCORE, ELEMENTARY AND SECONDARY GRADES, BY CLASS SIZE INTERVALS

	Elementary (Grades 3, 4, 5, 6) Total N 2106		Secondary (Grades 10, 11, 12) Total N 2181	
Number Students in Class	N	Mean Difference Score	N	Mean Difference Score
1- 5	14	10.00	16	6.23
6-10	34	10.09	162	8.90
11-15	71	10.04	351	7.66
16-20	376	8.72	566	4.51
21-25	999	8.18	553	4.55
26-30	494	6.89	320	4.51
31-35	69	6.60	74	3.99
36-40	10	9.10	37	5.65
41-50	15	4.70	32	6.13
over 50	14	2.07	64	4.91

IAR-

Research Bulletin

INSTITUTE OF ADMINISTRATIVE RESEARCH
Teachers College • Columbia University

Volume 9, No. 1 November, 1968

Devoted to research carried on by the Institute of Administrative Research and related groups, review of results, report of work in progress, discussion of theory and design, and implications for educational policies

Published in November, February and May
by the Institute of Administrative Research

•

WILLIAM S. VINCENT, Director
ERNST AUERBACHER, Research Associate
PAUL R. MORT FELLOW
DAVID L. DONOVAN
U. S. OFFICE OF EDUCATION RESEARCH TRAINEES
E. Robert Bagley
JOHN J. CASEY
ANHONY M. CROSSWELL
HARBISON POOL
RESEARCH FELLOWS
ROBERT K. LEDFORD
MARTIN N. OLSON
STAN SILVER
BRIAN SIMPSON
JOHN D. VANDEN BRINK

•

Subscription rate: $2.00 per year

Copyright 1968: May be quoted without permission provided source is named

NON-PROFIT ORG.
U. S. POSTAGE
PAID
WALDEN, N. Y.
PERMIT NO. 79

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY William S. Vincent Director TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."